Group 2
Siva Satya Sri Ganesh Seeram,
Hansi Abeynanda,
Eren Berk Kama,
Irshad Ahmed Meer,
Zinat Behdad

EP3260: Machine Learning Over Networks
Homework Assignment 2
Due Date: February 15, 2023

Problem 2.1

Consider Human Activity Recognition Using Smartphones dataset $\{(\mathbf{x}_i, y_i)\}_{i\in[N]}$, with inputs defined as the accelerometer and gyroscope sensors, and outputs defined as moving (e.g., walking , running, dancing) or not (sitting or standing). Consider the logistic ridge regression loss function

$$\underset{\mathbf{w}}{\text{minimize}} \quad f(\mathbf{w}) = \frac{1}{N}\sum_{i\in[N]} f_i(\mathbf{w}) + \lambda\|\mathbf{w}\|_2^2,$$

where $f_i(\mathbf{w}) = \log\left(1+\exp\{-y_i\mathbf{w}^T\mathbf{x}_i\}\right)$.

Then, address the following questions:

(a) Is f Lipschitz continuous? If so, find a small B?

(b) Is f_i smooth? If so, find a small L for f_i? What about f?

(c) Is f strongly convex? If so, find a high μ?

complete

(a) f is Lipschitz continuous if

Lipschitz continuity (bounded gradients)

$$\|\boldsymbol{w}\|_2 \le D \Rightarrow \|\nabla f(\boldsymbol{w})\|_2 \le B$$

I or $\|\boldsymbol{w}_1\|_2, \|\boldsymbol{w}_2\|_2 \le D \Rightarrow |f(\boldsymbol{w}_2) - f(\boldsymbol{w}_1)| \le B\|\boldsymbol{w}_2 - \boldsymbol{w}_1\|_2$

$$\nabla f(w) = \frac{1}{N}\sum_{i\in[N]} \nabla f_i(w) + 2\lambda w$$

$$\nabla f_i(w) = \frac{-y_i x_i \, e^{-y_i w^T x_i}}{1+e^{-y_i w^T x_i}} = \frac{-y_i x_i}{1+e^{y_i w^T x_i}} < e$$

$$\|\nabla f_i(w)\| = \left\| \frac{-y_i x_i}{1+e^{y_i w^T x_i}} \right\| \le |y_i| \cdot \|x_i\|$$

$$\|\nabla f(w)\|_2 = \left\| \frac{1}{N} \sum_{i\in[N]} \nabla f_i(w) + 2\lambda w \right\|_2 \le B$$

$$\le \frac{1}{N} \sum_{i\in[N]} \underbrace{\|\nabla f_i(w)\|}_{\le |y_i|\cdot\|x_i\|} + 2\lambda \underbrace{\|w\|}_{\le D}$$

$$\le \underbrace{\frac{1}{N} \sum |y_i| \cdot \|x_i\| + 2\lambda D}_{B}$$

(b) $\| \nabla f_i(w_2) - \nabla f_i(w_1) \|_2 =$

$$= \left\| \frac{-y_i x_i e^{-y_i w_2^T x_i}}{1+e^{-y_i w_2^T x_i}} - \frac{-y_i x_i e^{-y_i w_1^T x_i}}{1+e^{-y_i w_1^T x_i}} \right\|$$

$$= \|y_i\| \cdot \|x_i\| \left\| \frac{1}{1+e^{y_i w_2^T x_i}} - \frac{1}{1+e^{y_i w_1^T x_i}} \right\| \quad (1)$$

Let $h(w) = \dfrac{1}{1+e^{y_i w^T x_i}}$

$h(w)$ is continuous and differentiable

$\therefore$ By mean value theorem, $\exists\, z \in (w_1, w_2)$ s.t

$h(w_2) - h(w_1) = \nabla h(z)^T (w_2 - w_1)$

$\| h(w_2) - h(w_1) \| \leq \| \nabla h(z) \| \cdot \| w_2 - w_1 \| \quad (2)$

where $\nabla h(w) = \dfrac{-y_i x_i e^{y_i w^T x_i}}{(1+e^{y_i w^T x_i})^2}$

$\therefore (2) \Rightarrow$

$$\| h(w_2) - h(w_1) \| \leq \| y_i x_i \| \underbrace{\left\| \frac{e^{y_i z^T x_i}}{(1+e^{y_i z^T x_i})^2} \right\|}_{\leq 1} \cdot \| w_2 - w_1 \|$$

$$\leq |y_i| \, \|x_i\| \, \|w_2 - w_1\|$$

$\therefore$ (1) $\Longrightarrow$

$$\|\nabla f_i(w_2) - \nabla f_i(w_1)\| = |y_i| \, \|x_i\| \, \|h(w_2) - h(w_1)\|$$

$$\leq \underbrace{|y_i|^2 \cdot \|x_i\|^2}_{L} \|w_2 - w_1\|$$

$\therefore$ a small L for f_i is $|y_i|^2 \cdot \|x_i\|^2$

About f:

∇f is also Lipschitz continuous

Proof:

$$\|\nabla f(w_2) - \nabla f(w_1)\| =$$

$$\left\| \frac{1}{N} \sum_{i \in [N]} \left(\nabla f_i(w_2) - \nabla f_i(w_1)\right) + 2\lambda (w_2 - w_1) \right\|$$

$$\leq \frac{1}{N} \sum_{i \in [N]} \|\nabla f_i(w_2) - \nabla f_i(w_1)\| + 2\lambda \|w_2 - w_1\|$$

$$\leq \underbrace{\left[\frac{1}{N} \sum_{i \in [N]} |y_i|^2 \cdot \|x_i\|^2 + 2\lambda\right] \|w_2 - w_1\|}_{L_1}$$

$\Longrightarrow f$ is L_1-smooth

(C) f is strongly convex, because
$\lambda \|w\|^2$ is strongly convex and
$\frac{1}{N} \sum_{i \in [N]} f_i(w)$ is convex.

Let $g(w) = \lambda \|w\|^2 = \lambda w^T w$

$\nabla g(w) = 2\lambda w \implies \nabla^2 g(w) = 2\lambda I$

$\implies g$ is strongly convex, with $\boxed{\mu = 2\lambda}$

$\therefore f$ is also 2λ-strongly convex